海子水牛瘤胃微生物的宏基因组学分析

张慧敏　夏海磊　黄　强　绪　欣　毛永江　岑　宁　杨章平*

(扬州大学动物科学与技术学院，扬州 225009)

摘　要：为系统探讨海子水牛瘤胃内的微生物组成及木质纤维素降解酶系，本试验利用基于高通量测序的宏基因组学技术对海子水牛（2.5 岁左右，平均体重为 493 kg）瘤胃液样本进行组学分析。结果表明，共获得了 77 283 638 条 reads，并拼接出 744 712 个 scaffold。经 prodigal 分析后，共预测出 827 044 个基因。通过基因注释发现海子水牛瘤胃中含有多种木质纤维素降解微生物，如生黄瘤胃球菌（*Ruminococcus flavefaciens*）、白色瘤胃球菌（*Ruminococcus albus*）、产琥珀酸丝状杆菌（*Fibrobacter succinogenes*）及栖瘤胃普雷沃氏菌（*Prevotella ruminicola*）。此外，还发现有 38 011 个基因编码蛋白具有木质纤维素降解酶活性，其中糖苷水解酶（*GH*）基因数量最多（17 877 个），糖基转移酶（*GT*）（8 637 个）、碳水化合物结合模块（*CBM*）（4 693 个）和碳水化合物酯酶（*CE*）基因（4 214 个）次之，多糖裂合酶（PL）（1 296 个）和辅助氧化还原酶（AA）基因（934 个）较少。在 *GH* 基因中，归属于 *GH*2、*GH*43、*GH*97、*GH*3 家族的基因较多，且编码蛋白具有寡糖降解酶活性的基因数量最多。此外还发现了少量的纤维小体组分蛋白基因。结合其他物种肠胃宏基因组中 *GH* 基因比对分析，发现海子水牛瘤胃中的纤维素酶、半纤维素酶和分支降解酶比例与奶牛瘤胃较为接近。由此可见，海子水牛瘤胃内含有丰富的木质纤维素降解微生物及酶系，这将为筛选具有工业应用潜力的酶基因奠定理论基础。

关键词：海子水牛；瘤胃；宏基因组；微生物；木质纤维素降解酶

中图分类号：S823

木质纤维素是含有纤维素、半纤维素、木质素等成分的植物生物质的统称[1]。瘤胃是自然界中降解木质纤维素能力最强的生态系统之一，能够在48 h内降解饲粮中60%~65%的木质

收稿日期：2017-05-04

基金项目：国家自然科学基金青年基金（31702142）；江苏省自然科学基金青年基金（BK20160455）；江苏省高校自然科学研究面上项目资助（15KJB230004）；扬州大学大学生科技创新基金资助（x20160673）

作者简介：张慧敏（1986-），女，山东潍坊人，讲师，从事反刍动物饲料酶制剂研究。E-mail: minmin-911@163.com

*通信作者：杨章平，教授，博士生导师，E-mail: yzp@yzu.edu.cn

纤维素，为反刍动物提供营养物质，可见瘤胃内蕴藏着丰富的木质纤维素降解微生物和酶系[2]。瘤胃内微生物种类繁多、数量巨大，共有超过3 000种微生物，既包括细菌、原虫、真菌和古细菌，也有噬菌体和病毒[3]。瘤胃是一座巨大的生物资源库，蕴藏着许多有具有工业应用潜力的木质纤维素降解酶系。Zhou等[4]从牦牛瘤胃中发现了一个双功能酶，同时具有葡萄糖苷酶及木糖苷酶的活性。Huang等[5]从湖羊瘤胃中获得了一种嗜低温的木聚糖酶基因。但瘤胃是一个严格厌氧的环境，绝大多数瘤胃微生物是不可培养的，这就大大限制了人们对瘤胃基因资源的探索研究。随着高通量测序技术的发展，宏基因组测序已功运用于分析特定环境中的微生物组成、代谢通路研究以及酶蛋白的功能分析等[6]。

目前，已有许多应用高通量测序技术研究反刍动物瘤胃微生物的报道。Hess等[7]在2011年首次对奶牛瘤胃微生物的宏基因组进行了测序，共发现了27 755个木质纤维素降解酶基因，其中糖苷水解酶（glycoside hydrolase，GH）家族GH13（3 442个）和GH3（2 844个）的基因最多，而含有黏连模块和对接模块的基因分别为80和188个，这些基因为发掘新型木质纤维素降解酶提供了可能性。Patel等[8]利用青草、干草饲喂水牛，采样后进行宏基因组高通量测序，结果发现青草组的水牛瘤胃中GH28及果胶裂解酶的基因数量显著多于干草组。Singh等[9]利用高通量测序技术对印度水牛瘤胃微生物宏基因组进行测序，共获得了137 270个重叠群（contigs），预测其中2 614个具有降解木质纤维素的功能。王继文等[10]对山羊瘤胃与粪便微生物基因组进行了比对研究，发现二者存在较大差异，瘤胃微生物主要以拟杆菌门为主，厚壁菌门次之，相对丰度最高的属为普雷沃菌属，而粪便中的微生物以厚壁菌门为主，拟杆菌门次之，相对丰度最高的属为瘤胃球菌科下未分类的属。

海子水牛属沼泽型水牛，分布于江苏北部沿海地区，是我国主要的地方良种之一，它适应当地的自然生态条件，抗逆性突出，耐粗饲，在草质较差的情况下，仍能较好地生长及劳役，可见海子水牛瘤胃内蕴藏着独特的木质纤维素降解微生物和酶系。然而，到目前为止，对于沼泽型水牛瘤胃内的微生物组成及其功能基因尚鲜见报道。本研究借助于高通量测序技术，首次获得海子水牛瘤胃微生物的宏基因组数据，经数据库比对分析后，获得功能注释，结合其他物种肠胃宏基因组信息进行比对研究，旨在系统探讨海子水牛瘤胃内的微生物组成及木质纤维素降解酶系，以期从中发掘新型具有工业应用潜力的木质纤维素降解酶。

1 材料与方法

1.1 样品采集

选取3头2.5岁左右的海子水牛，平均体重为493 kg。试验水牛的饲养管理条件一致，每天饲喂2次，自由采食，饲粮组成及营养水平见表1。饲喂60 d后进行称重并屠宰，屠宰后立刻采集瘤胃液样品，经4层灭菌纱布过滤后，迅速置于液氮中。

在相同的饲粮及生活环境下，牛瘤胃内的微生物群落组成较为相似[7,11]，因此我们随机选取1头海子水牛的瘤胃液进行下一步测序分析。

表1 饲粮组成及营养水平（风干基础）

Table 1 Composition and nutrient levels of the diet (air-dry basis) %

原料 Ingredients	含量 Content	营养水平 Nutrient levels	含量 Content
青贮玉米 Silage corn	30.00	干物质 DM	96.06
花生秧 Peanut vine	10.00	粗蛋白质 CP	9.80
稻草 Rice straw	45.00	中性洗涤纤维 NDF	50.23
玉米 Corn	7.00	酸性洗涤纤维 ADF	42.46
米糠 Rice bran	2.50	有机物 OM	87.98
豆粕 Soybean meal	3.00		
小麦麸 Wheat bran	2.50		
合计 Total	100.00		

1.2 宏基因组DNA的提取

按照QIAamp DNA micro kit （德国Qiagen）试剂盒说明书的要求，进行瘤胃液宏基因组DNA的分离和纯化。使用1.0%的琼脂糖凝胶电泳和Nano Drop分光光度计，对分离纯化后的DNA进行检测和评价。

1.3 宏基因组测序及分析

纯化后的样品送上海欧易生物公司进行高通量测序。首先利用Covaris M220进行宏基因组DNA片段化处理，打断至300 bp左右，末端连接Y字形接头，通过接头PCR将DNA片段连接到芯片上形成“桥”并制成文库，随后通过Illumina HiSeq 2000高通量测序平台进行测序。

获得测序数据后，首先进行宿主基因的过滤和质量剪切等优化处理；其次使用软件SOAP denovo 2.04 对测序数据进行拼接组装，组装时选用的 kmer 参数为 65；然后使用软件prodigal 2.6.3 对基因组序列进行基因预测；最后进行数据库比对分析。

通过 BLAST 将预测得到的基因序列与 GenBank 的非冗余核苷酸序列数据库

（Nonredundant database，NR）进行比对（ftp://ftp.ncbi.nih.gov/pub/taxonomy），将 evalue≤1e-5 的序列作为有意义的序列，并获得物种注释信息；最后将预测基因与 COG 数据库（ftp://ftp.ncbi.nih.gov/pub/COG/COG/kyva）、GO 数据库（http://www.geneontology.org/）和 KEGG 数据库（http://www.genome.jp/kegg/pathway.html）进行比对，获得基因功能注释。

1.4　碳水化合物基因分析

使用 CAZy 数据库（http://www.cazy.org/）的对应工具 hmmscan（http://hmmer.janelia.org/search/hmmscan）将预测基因与CAZy数据库进行比对，获得基因对应的碳水化合物活性酶注释信息，然后使用碳水化合物活性酶对应的基因丰度总和计算该碳水化合物活性酶的丰度。根据蛋白质结构域中氨基酸序列的相似性，可以将不同物种来源的碳水化合物活性酶分成GH、糖基转移酶（glycosyl transferase，GT）、多糖裂合酶（polysaccharide lyase，PL）、碳水化合物酯酶（carbohydrate esterase，CE）、碳水化合物结合模块（carbohydrate-binding module，CBM）、辅助氧化还原酶（auxiliary activity，AA）共6大类蛋白质家族。同时将海子水牛的CAZy分析结果，与奶牛及印度水牛瘤胃、白蚁后肠及鸡盲肠宏基因组中的CAZy分析结果进行比对。

2　结果与分析

2.1　宏基因组测序结果

经 Illumina Hiseq 高通量测序平台测序，获得 77 283 638 条 reads，过滤水牛宿主基因（GCA_000003055.5,Bostaurus Bos_taurus_UMD_3.1.1）后共获得了 10.79 GB 的测序数据。使用软件 SOAP denovo 2.04 对测序数据进行拼接组装，共得到 744 712 个 scaffold，约 250 Mbp 的宏基因组序列（表 2）。

表 2　组装结果统计表

Table 2　The result of the assembly

项目 Items	Scaffold
总数 Total number/个	744 712
总长度 Total length/bp	250 094 761
Gap 数量 Gap No./个	1 777 360
平均长度 Average length/bp	336
N50 长度 N50 length/bp	302
N90 长度 N90 length/bp	210
最长长度 Maximum length/bp	185 379

最短长度 Minimum length/bp	200
GC 含量 GC content/%	46.14

使用软件 prodigal 2.6.3 对基因组序列进行基因预测，预测得到氨基酸序列共 827 044 条，基因平均长度 274 bp。

2.2　基因注释结果

将 827 044 条基因序列与 NR 数据库进行 BLAST 比对，选取 evalue≤1e-5 的序列，共 498 644 条。如图 1 所示，与 NR 数据库中序列的相似度主要分布在 50%以上，只有 7.8%的序列与 NR 数据库序列高度相似（相似度>95%），表明此次测序结果中含有许多新基因。

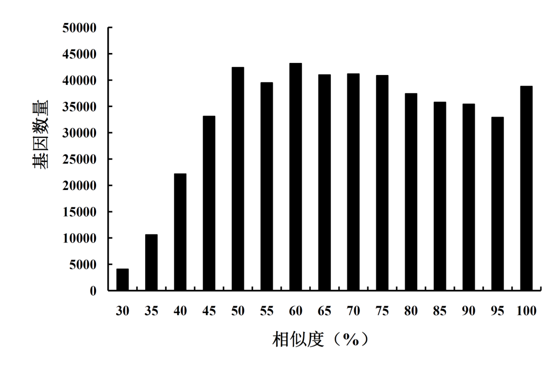
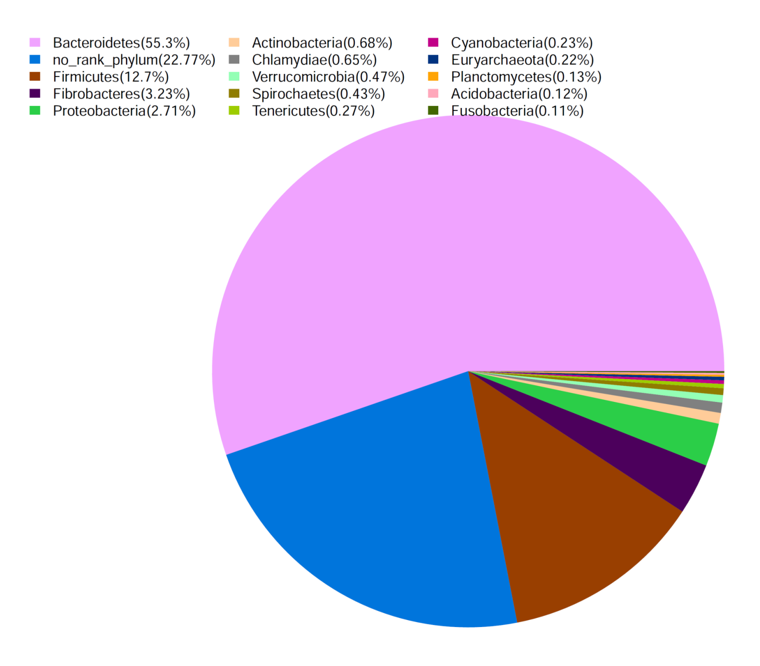


图 1　与 NR 数据库序列的相似度

Fig.1　Similarity with sequences in NR database

将测序数据与数据库比对，获得分析序列的蛋白功能注释信息。同时统计基因比对到各个物种的比例，结果显示蛋白编码基因来源于拟杆菌门（55.3%）的最多，来源于厚壁菌门的比例可达到 12.7%,3.23%及 2.71%的基因分别来源于纤维杆菌门及变形菌门（图 2）。



不同的颜色代表不同的门，饼面积的大小代表注释到该门的基因的比例。

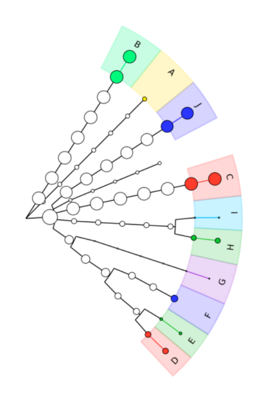
Different colors represent different phyla, and the area shows the proportion of gene classified to the phylum.

图 2　最多的 15 个门的分布

Fig.2　Distribution of the top 15 phyla

此外，将样品中物种水平上的分类层级关系绘制成进化分支图（图 3），从根部开始向外，依次为界-门-纲-目-科-属-种-菌株的分类。共分类出来了 3 个界：细菌界、古菌界、病毒界，海子水牛瘤胃中的微生物大部分为细菌界。在菌株的分类中，未分类的甲烷短杆菌(*Methanobrevibacter* unclassified)及猪沙波病毒(*Porcine sapelovirus*)分别属于古菌界及病毒界。而细菌界中共测得 8 种主要菌株，分别为栖瘤胃普雷沃氏菌（*Prevotella ruminicola*）、生黄瘤胃球菌（*Ruminococcus flavefaciens*）、白色瘤胃球菌（*Ruminococcus albus*）、未分类的丁酸弧菌（*Butyrivibrio* unclassified）、链球菌（*Streptococcus infantarius*）、密螺旋体 JC4（*Treponema* sp. JC4）及伯氏密螺旋体（*Treponema bryantii*）、产琥珀酸纤维杆菌（*Fibrobacter succinogenes*）。

批注 [微软用户1]: 排版
拟杆菌门 Bacteroidetes(55.3%)
未分类门 No_rank_phylum（22.77%）
厚壁菌门 Firmicutes（12.70%）
纤维杆菌门 Fibrobacteres（3.23%）
变形菌门 Proteobacteria（2.71%）
放线菌门 Actinobacteria（0.68%）
衣原体门 Chlamydiae（0.65%）
疣微菌门 Verrucomicrobia（0.47%）
螺旋体门 Spirochaetes（0.43%）
柔膜菌门 Tenericutes（0.27%）
蓝藻细菌 Cyanobacteria（0.23%）
广古菌门 Euryarchaeota（0.22%）
浮霉菌门 Planctomycetes（0.13%）
酸杆菌门 Acidobacteria（0.12%）
梭杆菌门 Fusobacteria（0.11%）



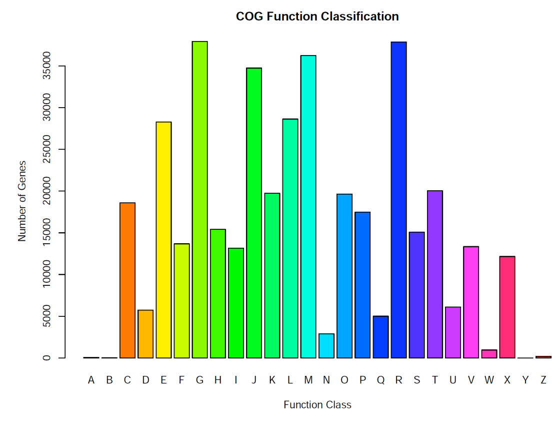
A:未分类的甲烷短杆菌 *Methanobrevibacter* unclassified；B:猪萨佩罗病毒 *Porcine sapelovirus*；C:栖瘤胃普雷沃氏菌 *Prevotella ruminicola*；D：生黄瘤胃球菌属 *Ruminococcus flavefaciens*；E:白色瘤胃球菌属 *Ruminococcus albus*；F:未分类的丁酸弧菌属 *Butyrivibrio* unclassified；G:链球菌 *Streptococcus infantarius*；H:梅毒螺旋体菌 JC4 *Treponema* sp. JC4；I:布赖恩蒂密螺旋体 *Treponema bryantii*；J:产琥珀酸纤维杆菌属 *Fibrobacter succinogenes*。

图 3　物种的分类

Fig.3　The classification of species

2.3　COG 数据库比对结果

对宏基因组中的预测基因进行 COG 功能分类预测。发现功能集中在碳水化合物转运及代谢的基因是最多的，数量达到 37 933 个，由此可见该宏基因组内可能含有大量的木质纤维素降解酶基因（图 4）。



A:RNA 加工和修改 RNAprocessingandmodification；B:染色质结构与动力学 Chromatinstructureanddynamics；C:能量产生和转换 Energyproductionandconversion；D:细胞周期控制、细胞分裂、染色体分区 Cellcyclecontrol,celldivision,chromosomepartitioning；E:氨基酸运输和代谢

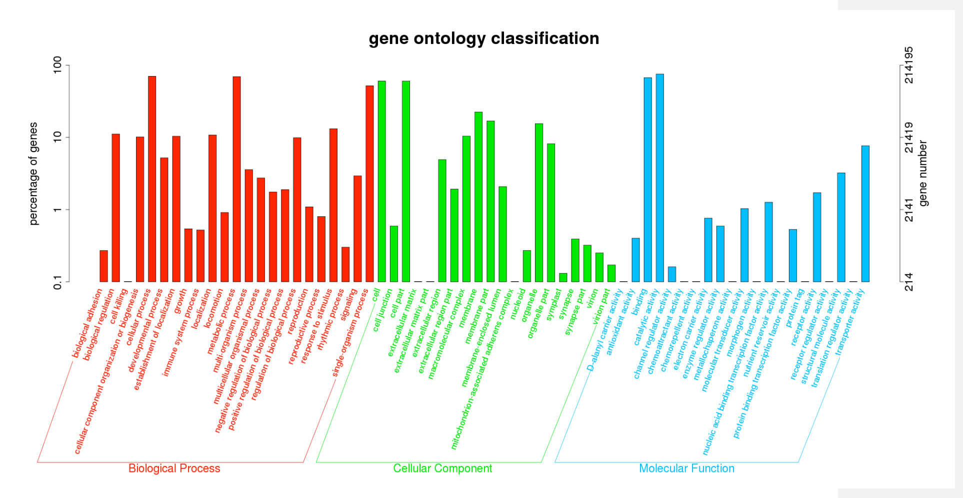
Aminoacidtransportandmetabolism；F: 核苷酸运输和代谢 Nucleotidetransportandmetabolism；G: 碳水化合物的运输和代谢 Carbohydratetransportandmetabolism；H:辅酶运输和代谢 Coenzymetransportandmetabolism；I:脂质运输和代谢 Lipidtransportandmetabolism；J:翻译,核糖体结构和生物转化 Translation,ribosomalstructureandbiogenesis；K:转录 Transcription；L:复制、重组和修复 Replication,recombinationandrepair；M:细胞壁/膜/包膜生物合成 Cellwall/membrane/envelopebiogenesis；N:细胞活性 Cellmotility；O:蛋白质转译后的修改,蛋白质转化,分子伴侣 Posttranslationalmodification,proteinturnover,chaperones；P: 无机离子运输和代谢 Inorganiciontransportandmetabolism；Q: 次生代谢产物生物合成、运输和分解代谢 Secondarymetabolitesbiosynthesis,transportandcatabolism；R:通用功能预测 Generalfunctionpredictiononly S:未知功能 Functionunknown；T:信号转导机制 Signaltransductionmechanisms；U:胞内运输、分泌和膜泡运输 Intracellulartrafficking,secretion,andvesiculartransport；V:防御机制 Defensemechanisms；W:细胞外结构 Extracellularstructures；X:移动基因组:噬菌体原,转座子 Mobilome:prophages,transposons；Y:核结构 Nuclearstructure；Z:细胞骨架 Cytoskeleton。

图 4　COG 分类

Fig.4　COG classification

2.4　GO 分类结果

对得到的基因进行 GO 分类，统计基因在生物过程（biological process，BP）、细胞成分（cellular component，CC）和分子功能（molecular function，MF）3 个 GO 类别中。在 BP 类中，集中在代谢、细胞外加工的基因最多；在 CC 类中，集中在细胞、细胞组分的基因最多；在 MF 类中，集中在结合以及催化功能的基因最多（图 5）。



A1:生物黏附 biological adhesion;A2:生物调节 biological regulation;A3:细胞死亡 cell killing;A4：组织或生物合成细胞组件 cellular component organization or biogenesis;A5：细胞过程 cellular process;A6:发育过程 developmental process;A7：定位建立 establishment of localization;A8：增长 growth;A9：免疫系统的过程 immune system process;A10：定位 localization;A11:移动 locomotion;A12:代谢过程 metabolic process;A13:

多细胞生物的过程 multicellular organismal process;A14:多组织过程 multi-organism process;A15:生物过程反向调节 negative regulation of biological process;A16:生物过程正向调节 positive regulation of biological process;A17:生物过程的调节 regulation of biological process;A18:繁殖 reproduction;A19:生殖过程 reproductive process;A20:刺激反应 response to stimulus;A21：节律过程 rhythmic process;A22：信号 signaling;A23：单组织过程 single-organism process;B1：细胞 cell;B2：细胞连接 cell junction;B3：细胞部分 cell part;B4：细胞外基质 extracellular matrix;B5：细胞外基质部分 extracellular matrix part;B6：细胞外区域 extracellular region;B7：细胞外区域部分 extracellular region part;B8：大分子复合物 macromolecular complex;B9：膜 membrane;B10：膜部分 membrane part;B11：膜封闭腔 membrane-enclosed lumen;B12：线粒体相关黏附复合体 mitochondrion-associated adherens complex;B13：类核 nucleoid;B14：细胞器 organelle;B15:细胞器的部分 organelle part;B16:共质体 symplast;B17:突触 synapse;B18:突触部分 synapse part;B19 病毒粒子 virion;B20：病毒粒子部分 virion part;C1：*D*-丙氨酰载体功能 *D*-alanyl carrier activity;C2:抗氧化功能 antioxidant activity;C3：结合 binding;C4：催化功能 catalytic activity;C5：通道调节功能 channel regulator activity;C6：化学引诱物功能 chemoattractant activity;C7:化学排斥物功能 chemorepellent activity;C8:电子载体功能 electron carrier activity;C9：酶调节功能 enzyme regulator activity;C10：金属伴侣功能 metallochaperone activity;C11：分子传感器功能 molecular transducer activity;C12:成形素功能 morphogen activity;C13：核酸结合转录因子功能 nucleic acid binding transcription factor activity;C14：营养储存功能 nutrient reservoir activity;C15：蛋白结合转录因子功能 protein binding transcription factor activity;C16：蛋白质标记 protein tag;C17：受体功能 receptor activity;C18:受体调节功能 receptor regulator activity;C19：结构分子功能 structural molecule activity;C20：翻译调节功能 translation regulator activity;C21：运输功能 transporter activity。

A 代表注释到 BP 的条目，B 代表注释到 CC 的条目，C 代表注释到 MF 的条目，横坐标为条目名称，纵坐标表示条目的数量或比例。

A represents the term of BP, B represents the term of CC, and C represents the term of MF. The abscissa was term name, and the ordinate was the number or percentage of term.

图 5　最多的 20 个 CO 条目分类

Fig.5　Top 20 GO terms classification

2.5　KEGG 分析结果

105 560 个基因被富集到 247 条 KEGG 通路中，表 3 为富集最明显的前 10 条 KEGG 通路。大部分基因被富集在小分子的代谢中，如氨基酸的生物合成、嘌呤代谢、嘧啶代谢、碳代谢、氨酰 tRNA 的生物合成等。此外，还有 5 236 个基因富集在淀粉和蔗糖代谢中。

表 3　富集最明显的 10 条 KEGG 通路

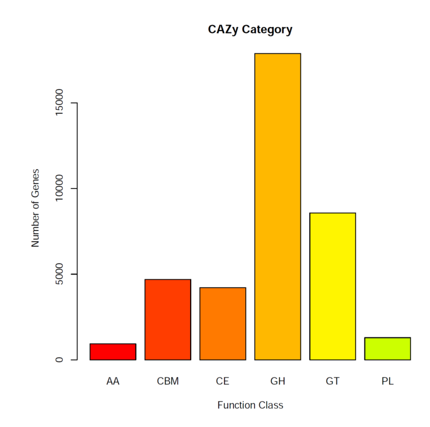
Table 3　Top 10 obviously gathered KEGG pathways

通路 Pathways	描述 Description	基因数量 Gene No./个
ko01230	氨基酸的生物合成	13 393
ko00230	嘌呤代谢	10 717
ko01200	碳代谢	10 664

ko00240	嘧啶代谢	9 467
ko00970	氨酰 tRNA 的生物合成	7 106
ko00500	淀粉和蔗糖代谢	5 236
ko00520	氨基糖和核苷酸糖代谢	5 200
ko00250	丙氨酸代谢	5 159
ko00010	糖酵解、糖原异生	4 969
ko03010	核糖体	4 872

2.6 CAZy 数据库比对结果

经 hmmscan 分析，发现有 38 011 个基因比对到 CAZy 数据库中，统计结果如图 6 所示，宏基因组中的 *GH* 基因数量最多（17 877 个），*GT*（8 637 个）、*CBM*（4 693 个）和 *CE* 基因（4 214 个）次之，*PL* 和 *AA* 基因较少，数量分别为 1 296 和 934 个。此外还发现了 42 个黏连蛋白（cohesin）基因、236 个对接蛋白（dockerin）基因以及 82 个 S 层同源结构域（S-layer homology，SLH）。



AA：辅助氧化还原酶；*CBM*：碳水化合物结合模块；*CE*：碳水化合物酯酶；*GH*：糖苷水解酶；*GT*：糖基转移酶；*PL*：多糖裂合酶。

AA: auxiliary activity *CBM*: carbohydrate binding modules; *CE*: carbohydrate esterases; *GH*: glycosyl hydrolases; *GT*: glycosyl transferases; *PL*: polysaccharide lyases.

图 6 CAZy 数据库比对结果

Fig.6 Distribution of CAZyme encoding genes

17 877 个 *GH* 基因分布在不同的 *GH* 基因家族，其中 *GH*2、*GH*43、*GH*97、*GH*3 家族的基因较多，*GH*12 和 *GH*62 家族的基因最少（表 4）。瘤胃中的木质纤维素降解酶可水解植

物细胞壁中的纤维素和半纤维素，根据水解底物类型的不同，可将海子水牛瘤胃宏基因组内的 *GH* 基因家族分类（表 5），有 932 个基因编码的蛋白质具备纤维素酶活性，其中 *GH*5 及 *GH*9 基因家族的数量较多；1 011 个基因编码的蛋白具备半纤维素酶活性，其中 *GH*10 基因家族的基因数量最多；分支降解酶（*GH*67 和 *GH*78）的编码基因共有 538 个；3 266 个基因编码的蛋白具有寡糖降解酶活性，其中 *GH*2 基因家族的基因数量最多。

表 4 海子水牛瘤胃微生物宏基因组中 *GH* 基因的分布

Table 4 Distribution of *GH* genes in *Haizi* buffalo rumen microorganism metagenome

项目 Items	数量 No./个	项目 Items	数量 No./个	项目 Items	数量 No./个
糖苷水解酶 1 *GH*1	31	糖苷水解酶 47 *GH*47	8	糖苷水解酶 93 *GH*93	27
糖苷水解酶 2 *GH*2	1 732	糖苷水解酶 48 *GH*48	6	糖苷水解酶 94 *GH*94	290
糖苷水解酶 3 *GH*3	860	糖苷水解酶 49 *GH*49	6	糖苷水解酶 95 *GH*95	547
糖苷水解酶 4 *GH*4	105	糖苷水解酶 50 *GH*50	99	糖苷水解酶 96 *GH*96	13
糖苷水解酶 5 *GH*5	515	糖苷水解酶 51 *GH*51	396	糖苷水解酶 97 *GH*97	940
糖苷水解酶 6 *GH*6	15	糖苷水解酶 52 *GH*52	7	糖苷水解酶 98 *GH*98	15
糖苷水解酶 7 *GH*7	9	糖苷水解酶 53 *GH*53	158	糖苷水解酶 99 *GH*99	42
糖苷水解酶 8 *GH*8	130	糖苷水解酶 54 *GH*54	54	糖苷水解酶 100 *GH*100	8
糖苷水解酶 9 *GH*9	337	糖苷水解酶 55 *GH*55	61	糖苷水解酶 101 *GH*101	7
糖苷水解酶 10 *GH*10	360	糖苷水解酶 56 *GH*56	9	糖苷水解酶 102 *GH*102	14
糖苷水解酶 11 *GH*11	56	糖苷水解酶 57 *GH*57	143	糖苷水解酶 103 *GH*103	18
糖苷水解酶 12 *GH*12	3	糖苷水解酶 58 *GH*58	5	糖苷水解酶 104 *GH*104	13
糖苷水解酶 13 *GH*13	679	糖苷水解酶 59 *GH*59	25	糖苷水解酶 105 *GH*105	204
糖苷水解酶 14 *GH*14	8	糖苷水解酶 62 *GH*62	3	糖苷水解酶 106 *GH*106	310
糖苷水解酶 15 *GH*15	11	糖苷水解酶 63 *GH*63	18	糖苷水解酶 107 *GH*107	13
糖苷水解酶 16 *GH*16	192	糖苷水解酶 64 *GH*64	8	糖苷水解酶 108 *GH*108	26
糖苷水解酶 17 *GH*17	13	糖苷水解酶 65 *GH*65	22	糖苷水解酶 109 *GH*109	315
糖苷水解酶 18 *GH*18	119	糖苷水解酶 66 *GH*66	60	糖苷水解酶 110 *GH*110	61
糖苷水解酶 19 *GH*19	5	糖苷水解酶 67 *GH*67	115	糖苷水解酶 111 *GH*111	5
糖苷水解酶 20 *GH*20	401	糖苷水解酶 68 *GH*68	6	糖苷水解酶 112 *GH*112	10
糖苷水解酶 22 *GH*22	6	糖苷水解酶 70 *GH*70	15	糖苷水解酶 113 *GH*113	18
糖苷水解酶 23 *GH*23	266	糖苷水解酶 71 *GH*71	9	糖苷水解酶 114 *GH*114	20
糖苷水解酶 24 *GH*24	68	糖苷水解酶 72 *GH*72	5	糖苷水解酶 115 *GH*115	320
糖苷水解酶 25 *GH*25	275	糖苷水解酶 73 *GH*73	187	糖苷水解酶 116 *GH*116	23
糖苷水解酶 26 *GH*26	163	糖苷水解酶 74 *GH*74	199	糖苷水解酶 117 *GH*117	21
糖苷水解酶 27 *GH*27	159	糖苷水解酶 75 *GH*75	4	糖苷水解酶 118 *GH*118	9
糖苷水解酶 28 *GH*28	302	糖苷水解酶 76 *GH*76	75	糖苷水解酶 116 *GH*119	201
糖苷水解酶 29 *GH*29	368	糖苷水解酶 77 *GH*77	310	糖苷水解酶 120 *GH*120	18
糖苷水解酶 30 *GH*30	192	糖苷水解酶 78 *GH*78	423	糖苷水解酶 121 *GH*121	25
糖苷水解酶 31 *GH*31	544	糖苷水解酶 79 *GH*79	11	糖苷水解酶 122 *GH*122	9
糖苷水解酶 32 *GH*32	189	糖苷水解酶 80 *GH*80	18	糖苷水解酶 123 *GH*123	50
糖苷水解酶 33 *GH*33	190	糖苷水解酶 81 *GH*81	4	糖苷水解酶 124 *GH*124	45

糖苷水解酶 34 *GH*34	12	糖苷水解酶 82 *GH*82	29	糖苷水解酶 125 *GH*125	52
糖苷水解酶 35 *GH*35	88	糖苷水解酶 83 *GH*83	52	糖苷水解酶 126 *GH*126	8
糖苷水解酶 36 *GH*36	218	糖苷水解酶 84 *GH*84	32	糖苷水解酶 127 *GH*127	258
糖苷水解酶 37 *GH*37	30	糖苷水解酶 85 *GH*85	19	糖苷水解酶 128 *GH*128	14
糖苷水解酶 38 *GH*38	47	糖苷水解酶 86 *GH*86	8	糖苷水解酶 129 *GH*129	31
糖苷水解酶 39 *GH*39	91	糖苷水解酶 87 *GH*87	24	糖苷水解酶 130 *GH*130	147
糖苷水解酶 42 *GH*42	49	糖苷水解酶 88 *GH*88	36	糖苷水解酶 131 *GH*131	7
糖苷水解酶 43 *GH*43	1 105	糖苷水解酶 89 *GH*89	199	糖苷水解酶 132 *GH*132	4
糖苷水解酶 44 *GH*44	28	糖苷水解酶 90 *GH*90	9	糖苷水解酶 133 *GH*133	167
糖苷水解酶 45 *GH*45	52	糖苷水解酶 91 *GH*91	13	糖苷水解酶 134 *GH*134	8
糖苷水解酶 46 *GH*46	16	糖苷水解酶 92 *GH*92	602	糖苷水解酶 135 *GH*135	13

表 5　不同动物的微生物宏基因组中 GH 基因的分布

Table 5　Distribution of GH genes in microorganism metagenome in different animals

项目 Items	海子水牛瘤胃 *Haizi* buffalo rumen	奶牛瘤胃 Dairy cow rumen[7]	印度水牛瘤胃 Indian buffalo rumen[9]	白蚁后肠 Termite hindgut[11]	鸡盲肠 Chicken cecum[12]
纤维素酶 Cellulases					
糖苷水解酶 5 *GH5*	2.88	3.85	1.50	5.71	0.19
糖苷水解酶 9 *GH9*	1.89	2.2	0.26	2.37	0.23
糖苷水解酶 44 *GH44*	0.16	0.07	0.00	0.32	0.00
糖苷水解酶 45 *GH45*	0.29	0.27	0.05	0.51	0.00
合计 Total	5.22	6.39	1.81	8.91	0.43
半纤维素酶 Hemicellulases					
糖苷水解酶 8 *GH8*	0.73	0.76	0.00	0.83	0.08
糖苷水解酶 10 *GH10*	2.01	1.65	2.64	5.39	0.62
糖苷水解酶 11 *GH11*	0.31	0.69	0.15	1.09	0.00
糖苷水解酶 26 *GH26*	0.91	0.76	1.03	1.28	0.27
糖苷水解酶 28 *GH28*	1.69	0.48	0.21	0.77	0.51
合计 Total	5.65	4.34	4.04	9.36	1.48
分支降解酶 Debranching enzymes					
糖苷水解酶 67 *GH67*	0.64	0.62	2.49	1.34	0.47
糖苷水解酶 78 *GH78*	2.37	3.85	2.85	0.44	2.84
合计 Total	3.01	4.47	5.34	1.78	3.31
寡糖降解酶 Oligosaccharide degrading enzymes					
糖苷水解酶 1 *GH1*	0.17	0.21	0.05	1.86	3.86
糖苷水解酶 2 *GH2*	9.69	11.68	9.85	4.36	12.24
糖苷水解酶 3 *GH3*	4.81	13.87	18.30	11.94	8.92
糖苷水解酶 29 *GH29*	2.06	2.06	2.28	0.89	2.14
糖苷水解酶 35 *GH35*	0.49	0.62	0.78	0.38	1.09
糖苷水解酶 38 *GH38*	0.26	0.89	0.26	2.63	1.21
糖苷水解酶 39 *GH39*	0.51	0.21	0.00	0.44	0.00
糖苷水解酶 42 *GH42*	0.27	0.27	0.10	2.76	0.94

合计 Total	18.26	29.81	31.62	25.28	30.4

3 讨 论

本试验通过高通量测序技术对海子水牛瘤胃液样本进行组学分析，共获得了 77 283 638 条 reads，并拼接出 744 712 个 Scaffold。将测序数据与 COG 数据库进行比对发现，功能集中在碳水化合物转运及代谢的基因是最多的，除此之外，富集在氨基酸运输及代谢、细胞壁及细胞膜生物合成基因也较多，这与印度水牛的报道相似[9]。在 GO 分类分析中发现，集中在结合以及催化功能的基因最多。由此可见，海子水牛瘤胃宏基因组内可能含有大量的木质纤维素降解酶基因。

根据基因注释到物种的比例可知，来源于拟杆菌门的基因较多，可占55.3%。Jami等[13]在奶牛瘤胃中也发现了相同的微生物分布。此外，根据物种分类结果可知，海子水牛瘤胃中的微生物大部分为细菌界，其中的生黄瘤胃球菌、白色瘤胃球菌及产琥珀酸丝状杆菌可产出大量的纤维素酶降解饲粮中的木质纤维素，被认为是瘤胃中降解木质纤维素的主要微生物，而栖瘤胃普雷沃氏菌可降解木聚糖及果胶[14-15]。

纤维素酶、半纤维素酶及果胶酶的协同作用可降解植物的细胞壁，为反刍动物提供营养物质，而这 3 类酶基因均存在于海子水牛的瘤胃中。*GH5* 及 *GH9* 基因家族的含量较高，二者可达到 *GH* 基因总量的 4.77%。GH10、GH11、GH26 和 GH28 家族的酶具有半纤维素酶活性，其中 GH10 和 GH11 家族属于木聚糖酶，它能从木聚糖主链的内部随机切割 β-1,4 木糖苷键，将其降解为低聚木糖和少量木糖，而 CE1、CE4、CE6 及 CE7 家族的木聚糖酯酶可作用于木聚糖的支链[8,16]。此外，PL1、PL9 及 PL10 家族的酶具有果胶裂解酶活性，而 CE8 家族的酶具有果胶甲基酯酶活性，可脱去果胶中的甲酯基团，生成果胶酸[17]。在 GH 分类中，寡糖降解酶的比例最高，可能是由于经过纤维素酶、半纤维素酶及果胶酶的协同作用后，植物细胞壁被降解为大量的低聚寡糖，这些物质需要被进一步降解后才可被反刍动物吸收利用。而寡糖降解酶中的 GH2 家族的酶数量最多，主要包括 β-*D*-半乳糖苷酶、β-葡萄糖醛酸酶、β-*D*-甘露糖苷酶及外切 β-氨基葡萄糖苷酶[18]。

此外，本试验还在海子水牛瘤胃内发现了 42 个黏连蛋白基因、236 个对接蛋白基因以及 82 个 S 层同源结构域，它们均是构成纤维小体的组分。纤维小体是一种由木质纤维素降解酶依靠脚手架蛋白形成的多酶复合体[19]，并借助黏附蛋白固定于细胞表面，可持续、高

效地降解木质纤维素。目前已在生黄瘤胃球菌和白色瘤胃球菌中发现了纤维小体[20]。Hess 等[7]发现奶牛瘤胃中含有黏连蛋白和对接蛋白的基因分别为 80 和 188 个。

将海子水牛的瘤胃宏基因组与奶牛[7]和印度水牛的瘤胃[9]、白蚁后肠[11]及鸡盲肠[12]宏基因组数据中的 *GH* 基因进行分类比对，发现在不同的宏基因组中，寡糖降解酶比例均为最高。但与其他 4 种动物肠胃微生物宏基因组相比，海子水牛中的寡糖降解酶数量最少。海子水牛和奶牛瘤胃宏基因组中的纤维素酶、半纤维素酶和分支降解酶比例接近。印度水牛中的纤维素酶最少，但寡糖降解酶最多。海子水牛瘤胃中纤维素酶及寡糖降解酶的比例与印度水牛瘤胃存在许多不同，这可能与 2 种水牛饲粮中粗饲料与精饲料的比例有关，印度水牛饲粮中粗饲料与精饲料的比例为 75:25[9]，而海子水牛的则为 85:15，粗饲料含量高可能导致瘤胃内存在高比例的纤维素酶和低比例的寡糖降解酶。但在 *GH* 基因家族分布中，2 种水牛瘤胃中 *GH*6、*GH*7、*GH*12、*GH*48 及 *GH*62 家族基因数量均较少[8]。白蚁后肠微生物宏基因组中的纤维素酶和半纤维素酶基因最多，其中 GH5（5.71%）和 GH10（5.39%）家族的基因比例最高，主要是由于木材中纤维素及半纤维素的比例高于牧草。鸡盲肠中的纤维素酶及半纤维素酶的比例最低，但寡糖降解酶比例较高。由此可见宿主基因型及饲料的不同可影响肠胃 *GH* 基因的组成。

4　结　论

①通过高通量测序发现海子水牛瘤胃中含有多种木质纤维素降解微生物，如生黄瘤胃球菌、白色瘤胃球菌、产琥珀酸丝状杆菌及栖瘤胃普雷沃氏菌。可见，海子水牛瘤胃内含有丰富的木质纤维素降解微生物。

②水牛瘤胃中还含有大量的糖苷水解酶基因，其中 *GH*2、*GH*43、*GH*97、*GH*3 家族的基因较多，且编码蛋白具有寡糖降解酶活性的基因数量最多。可见，海子水牛瘤胃内含有丰富的木质纤维素降解酶系。

参考文献：

[1]　邓玉营,黄振兴,阮文权,等.木质纤维素沼气体系中共培养菌群形成及适应性变化研究进展[J].应用与环境生物学报,2016,22(5):944–958.

[2]　WANG G R,DUAN Y L.Studies on lignocellulose degradation by rumen

microorganism[J].Advanced Materials Research,2014,853:253–259.
[3] KIM M,MORRISON M,YU Z T.Status of the phylogenetic diversity census of ruminal microbiomes[J].FEMS Microbiology Ecology,2011,76(1):49–63.
[4] ZHOU J,BAO L,CHANG L,et al.Beta-xylosidase activity of a GH3 glucosidase/xylosidase from yak rumen metagenome promotes the enzymatic degradation of hemicellulosic xylans[J].Letters in Applied Microbiology,2012,54(2):79–87.
[5] HUANG H Q,WANG G Z,ZHAO Y Y,et al.Direct and efficient cloning of full-length genes from environmental DNA by RT-qPCR and modified TAIL-PCR[J].Applied Microbiology and Biotechnology,2010,87(3):1141–1149.
[6] 王禄禄,王立志,周美丽.宏基因组学技术在反刍动物瘤胃微生态系统上的应用研究进展[J].中国微生态学杂志,2017,29(2):223–228.
[7] HESS M,SCZYRBA A,EGAN R,et al.Metagenomic discovery of biomass-degrading genes and genomes from cow rumen[J].Science,2011,331(6016):463–467.
[8] PATEL D D,PATEL A K,PARMAR N R,et al.Microbial and carbohydrate active enzyme profile of buffalo rumen metagenome and their alteration in response to variation in the diet[J].Gene,2014,545(1):88–94.
[9] SINGH K M,REDDY B,PATEL D,et al.High potential source for biomass degradation enzyme discovery and environmental aspects revealed through metagenomics of Indian buffalo rumen[J].BioMed Research International,2014,2014:267189.
[10] 王继文,王立志,闫天海,等.山羊瘤胃与粪便微生物多样性[J].动物营养学报,2015,27(8):2559–2571.
[11] WARNECKE F,LUGINBÜHL P,IVANOVA N,et al.Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite[J].Nature,2007,450(7169):560–565.
[12] QU A,BRULC J M,WILSON M K,et al.Comparative metagenomics reveals host specific metavirulomes and horizontal gene transfer elements in the chicken cecum microbiome[J].PLoS One,2008,3(8):e2945.
[13] JAMI E,MIZRAHI I.Composition and similarity of bovine rumen microbiota across

individual animals[J].PLoS One,2012,7(3):e33306.
[14] FLINT H J,BAYER E A.Plant cell wall breakdown by anaerobic microorganisms from the mammalian digestive tract[J].Annals of the New York Academy of Sciences,2008,1125(1):280–288.
[15] FLINT H J,BAYER E A,RINCON M T,et al.Polysaccharide utilization by gut bacteria:potential for new insights from genomic analysis[J].Nature Reviews Microbiology,2008,6(2):121–131.
[16] ZHANG H M,LI J F,WANG J Q,et al.Determinants for the improved thermostability of a mesophilic family 11 xylanase predicted by computational methods[J].Biotechnology for Biofuels,2014,7(1):3.
[17] YADAV S,YADAV P K,YADAV D,et al.Pectin lyase:a review[J].Process Biochemistry,2009,44(1):1–10.
[18] NASCIMENTO A S,MUNIZ J R C,APARÍCIO R,et al.Insights into the structure and function of fungal β-mannosidases from glycoside hydrolase family 2 based on multiple crystal structures of the *Trichoderma harzianum* enzyme[J].FEBS Journal,2014,281(18):4165–4178.
[19] HYEON J E,JEON S D,HAN S O.Cellulosome-based,*Clostridium*-derived multi-functional enzyme complexes for advanced biotechnology tool development:advances and applications[J].Biotechnology Advances,2013,31(6):936–944.
[20] DASSA B,BOROVOK I,RUIMY-ISRAELI V,et al.Rumen cellulosomics:divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six Ruminococcal strains[J].PLoS One,2014,9(7):e99221.

Metagenomic Analysis of Microorganisms in Rumen of *Haizi* Buffalo

ZHANG Huimin XIA Hailei HUANG Qiang XU Xin MAO Yongjiang CEN Ning YANG Zhangping*
（*College of Animal Science and Technology, Yangzhou University, Yangzhou* 225009, *China*）

*Corresponding author, professor, E-mail: yzp@yzu.edu.cn

(责任编辑 王智航)

Abstract: In order to investigate microorganism composition and lignocellulolytic enzymes in rumen of *Haizi* buffalo, rumen fluid sample of *Haizi* buffalo （about 2.5 years of age, average body weight was 493 kg） was surveyed using metagenomic high-throughput sequencing. The results showed as follws: a total of 77 283 638 reads were obtained, which were assembled into 744 712 scaffolds. And then, a total of 827 044 genes were predicted by prodigal analysis. Many microorganisms involved in lignocellulose degradation were identified by gene annotation, such as *Ruminococcus flavefaciens, Ruminococcus albus*, *Fibrobacter succinogenes* and *Prevotella ruminicola*. We identified potential 38 011 genes encoding lignocellulolytic enzymes including glycoside hydrolase (*GH*) (17 877 genes), glycosyl transferase (*GT*) (8 637 genes), carbohydrate binding module (*CBM*) (4 693 genes), carbohydrate esterase (*CE*) (4 214 genes), polysaccharide lyase (*PE*) (1 296 genes), and auxiliary activity (*AA*) （934 genes). Among GH genes, most genes belonged to *GH*2, *GH*43, *GH*97 and *GH*3 families, and a higher abundance of oligosaccharide degrading enzymes in buffalo rumen metagenome was identified. Besides that, we also predicted that some genes could encode the cellulosome component. Compared with *GH* genes in the gastrointestinal metagenome of other species, the proportions of cellulose, hemicellulase and debranching enzyme were similar between *Haizi* buffalo and dairy cow. These results demonstrate that buffalo rumen is considerably enriched in functional microorganisms and enzymes involved in lignocellulose degradation with great prospects to obtain enzyme genes that may be applied in the industry.

Key words: *Haizi* buffalo; rumen; metagenome; microorganism; lignocellulolytic enzyme